基于 TAR 技术应用法律领域的思考

——第六章读后感

作者在第六章内容中主要是介绍了 Technology-assisted review(TAR)在文件搜索中的应用，TAR 技术中最为主要的是 Supervised Machine Learning for TAR，其本质上是一种机器学习方法，即通过运用示例来训练程序从而达到让程序来划分文档的目的，从而减少操作人员的工作量，提高工作效率与工作精度。接下来笔者将从 TAR 技术的概念，优点以及劣势等方面讲述本章内容以及心得体会。

一、概念

作者将 TAR 技术的概念界定为"是使用计算机软件将集合中的每个文档分类为有响应或无响应的过程，或者根据人类对集合中文档的一小部分的审查和编码，将文档从最有可能有响应到最不可能有响应的顺序排列。"TAR 技术作为一种辅助性审查方法，其核心就在于从给定的集合中将符合需要的文档甄别出来，相比较繁杂的人工审查而言，TAR 技术更为简便和快捷，而且通过进行对比试验，也可以论证 TAR 技术往往能够保证更高的准确率与召回率（召回评估率）。作者在第三节内容中提到了目前主要采用的几种审查方法，Supervised Machine Learning for TAR 在文件审查中表现出了巨大的优越性，在 Supervised Machine Learning for TAR 中，主要可以分为 3 个模型，SPL、SAL 与 CAL。在这三种方法上，CAL 可以实现更高的自主性,也是作者最为推荐的一种文件审查方法，在进行一个初始集合的分析后，CAL 可以自主判断并且根据排序不断重新巩固和完善甄别规则，与 TAR1.0（SPL 和 SAL 相比，CAL 实现了更大程度的自主性，并且在甄别过程中，由于关于响应文档的新特征的不断确定与填充，CAL 可以实现更高的召回率与精确率，从这个层面来讲，无论是 TAR1.0 还是基于规则库的 TAR 技术，抑或是人工审查方法，都很难做到保证效率的前提下自动更新响应文档的特征。这就是 TAR2.0 的优势所在。

二、TAR 技术的优势

TAR（主要是指以 CAL 为代表的 TAR2.0）的优势可以总结为以下几点。首先是高自主性，开头已经说到，Supervised Machine Learning for TAR 实质上就是一种机器学习，操所人员通过程序设置与初始训练，让机器拥有自己的"意识"，并且根据一定的逻辑来甄别出符合要求的文件，高自主性可以将人从繁重且重复的筛选工作中解放出来，这将为法律工作者提供极大的便利，大多数情况下法律文件的筛选只涉及到案件的初始部分，但这一过程极其耗费精力与时间，TAR 的引入可以大大节省律师处理案件的时间，同时可以使得律师可以集中注意力去处理更为重要的事务。TAR 技术的第二个优点是高效率性，不同于人工审查的方式，TAR 技术在进行初始集合的训练后，可以提取出响应文档的特征，并且根据这一特征来判断新集合中的文档是否符合需求，从而从中进行挑选。无论是 TAR1.0 还是 TAR2.0，都遵循着这一工作流程，相比较人工审查而言，机器不需要像人工审查一样逐字逐句来确定法律文件是否符合要求，可以大幅度提升工作的效率与速度。众所周知，法律文件是及其繁杂的，在遵循判例法的国家中，对于先前判例的寻找是案件处理至为关键的一个步骤，搜索到的相关判例的数量将会直接影响裁判的质量，运用 TAR 技术将可以更好在互联网上寻找案件相关的法律文件，提升案件处理的效率。TAR 技术的第三个优势就是高精确率，在本章内容中，作

者运用统计学中的两个数值比率来反映 TAR 技术在文件搜索中所能起到的主要作用，即召回率与精确率。作者在第五小节内容中分别通过叙述“EDI study”、“TREC 会议的法律跟踪任务”以及“The 2011 Richmond Journal of Law and Technology Study”的例子来说明 TAR 技术在实践运用中所能达到的实际效果，以“The 2011 Richmond Journal of Law and Technology Study”为例，在本次的任务测试中，人工审查的召回率平均为 59%，准确率为 32%，F1 平均为 36%；TAR 的召回率平均为 77%，精确度为 85%，F1 平均为 80%。除了召回率数值没有统计学意义以外，在精确率与 F1 数值的测评上，TAR 技术全面占优。这就表明，即使在召回评估只能证明双方只是寻找到了一定比例文件的前提下，TAR 技术能够在搜索更少不相关文件的前提下完成指定的任务。这进一步证名了 TAR 技术在提高效率方面所发挥的巨大作用。高精确率意味着 TAR 技术可以在更小的偏差内找到符合需求的响应文档，并且可以减少人工的筛选量，从而高质量的完成搜索任务。最后一点，TAR 本身所体现的技术与传统工作融合恰恰对应了互联网时代下智能化的一种发展趋势。互联网时代与大数据时代的来临使得人工智能的发展上升到一个更高的平台，对于法律工作而言，传统的工作方式已经不能适应现时代的发展节奏，文本检索作为一项最为基础且最为繁杂的任务，其基本特征就是：工作基数大、专业化程度低。在传统的人工审查方式中， 对于相关性文件的审查是通过人工来进行的，这种审查方式不仅耗费时间，而且审查的范围也会因为各种因素的影响而有所局限。而互联网的发展则打破了地域与时间的桎梏，使得法律文件的搜索面更为广泛，搜索数量的激增伴随而来的就是搜索成本的上升，科技的进步成功的推动了传统领域的变革。就法律领域而言，近些年来，智慧法院、电子证据以及裁判文书上网等都体现了科技与法律的融合趋势，而 TAR 技术的运用可以更好与其相对应，从而推动整个法律领域智能化的全面改革。综合来看，TAR 技术的产生与应用完全符合时代的发展要求，而且 TAR 技术所展现出的优势也是其他审查方法所不可比拟的，其在法律领域的推广应用不仅具有合理性，而且有其必要性。

三、TAR 技术的劣势

TAR 技术虽然有着以上的众多的优势，但其本身并不是完美的。笔者在通读本章内容后，也产生了下列的疑惑和想法。首先，审查目标的不确定性。找到集合中的“所有仅相关“文档是所有审查工作的目标，作为文档是否为相关文档的标准，”所有仅相关“实质上不存在一个客观的定义，相关性的界定过程本身就是一个仁者见仁、智者见智的过程，没有一个具体的标准，对于审查目标界定的不确定性将会直接影响审查方法的评估。与统计学的计量标准不同的是，审查方法的目标的不确定性导致无法对于各种审查方法进行真正客观的评估，对于 TAR 技术而言更是如此，衡量 TAR 技术的数值标准，召回率与精确率都要求首先”所有仅相关“文件有一个准确数据，才能进行下阶段的评估，但基于上述原因，相关文件的确定不可能是一个客观的数值，在此基础上得出的召回率与准确率本身的说服力也将大打折扣。其次，TAR 技术是否能够全面反映所有相关性文件。作者在本章内容中提到在相关文档总数为 100 时，两组数据中不可能有超过 67 份共同的文档被两个审查者都认为是相关的。这个例子一方反映了相关性标准认定的主观性，另一方面，也说明即使给定的要求相同，不同的审查者基于自身对于相关性的认定也会得出不同的结论，在这种情况下，只要每个审阅者都是勤奋认真的工作，就没有根据说某一组的数据更为接近于理想状态，即每组数据都是正确的，只是评价标准不同。对于 TAR 技术而言，其自始至终都遵循的都是一套评

价标准，其认为的”所有且相关“文档是否全面也值得怀疑。第三，TAR 技术是否相较于人工审查全面占优。审查目标是尽可能找到更多的”所有仅相关“文档，”所有仅相关“的界定标准可以细分为两个标准，相关性与全面性。正如上文提到的，TAR 技术自始至终贯穿的是一套评价标准，其对于集合中相关文档的认定较为死板，标准较为单一，但是不可否认的是这种评价标准可以很大程度的提升工作效率。笔者认为，在集合文档数量不多的时候，人工审查方式可能会有更好的反馈结果。多个审查员基于自身对于相关性的认定会选出相对应的相关性文档，将这些文件进行汇总以后进行第二轮次的再筛选，重复筛选的过程将会持续多轮，最后将得到大多数审查员认同的文档认定为相关文档，这个过程相较于 TAR 的筛选过程而言会繁杂很多，但是其相关性标准的确定以及文件的筛选经过了多个审查员的评定，在这种情况下，笔者认为，最后确定下来的相关性文档会比 TAR 技术认定的文档更为准确以及全面。第四，TAR 技术市场化应用涉及的隐私保护问题，作者在本章内容中也提到，TAR 技术今后的发展目标除了使认定标准更为客观以及确定以外，更为重要的就是实现 TAR 技术的完全自动化，TAR 搜索的范围不再局限于人类审查者给定的集合，而是在更大的网络区域中搜索符合标准的”相关性文档‘，这种搜索方式固然可以更为全面收集符合标准的相关性文档，但是，它也有可能会在信息搜索过程中泄露网络用户的隐私。作者在本章开头已经说明了 TAR 技术涉及的只是电子发现某个具体的阶段，假设将 TAR 技术适用于更大的互联网区域，其技术本身并不知道如何甄别互联网上的信息属于隐私信息，而完全自动化的 TAR 排除人为干预，从文档选择到相关文档的确定的整个过程都由其自主完成，在这种条件设置下，如果不进行预先的人为审查环节以及后续的筛选环节，TAR 技术将会对互联网用户的信息产生极大的威胁性。同时，TAR 技术作为技术与法律结合的“新型产品”，目前尚未有明确的立法来对其进行规制，在这一区域基本上可以属于立法空白，而法律规制的缺失同样会反过来对于 TAR 技术的市场化应用产生阻碍作用。其实，对于网络用户隐私权的侵犯实质上只是有关 TAR 技术进行法律规制问题的一个缩影，对于一个新事物的出现，法律必须对于其可能导致的法律问题进行预估并且尽可能的作出应对方案，防止其对于社会产生不利影响。最后一点是 TAR 技术在我国是否有运用的必要性，这一点也是值得讨论。作者在本章内容中已经明确说明了 TAR 技术的目标与人工审查目标，在集合中尽可能全面的找到“所有仅相关文档”。TAR 技术在法律领域的应用就是在案件初期帮助律师或者是法官尽可能的找到更多与即将裁判案件具有相似性的生效裁判，从而辅助律师更有效的制定诉讼策略或者帮助法官作出正确的裁决。

四、TAR 技术在我国的适用空间

除了对于技术本身的探讨以外，我们也需要注意到 TAR 技术目前存在的窘境。根据两位作者的文章中所提到，目前只有少数国家采取了 TAR 技术，大多数国家中文件审查方式仍旧是以人工审查作为其主要的审查方式，TAR 技术虽然在科学实验中取得了巨大的成功，但是应用于社会生产中会产生怎样的问题尚未可知，究竟是否能够达到我们预想的结果也没有强有力的事实为其佐证。由此可见，TAR 技术对于完成从“实验室”到“市场”的转型尚需要一定的时间，而且社会民众对于 TAR 技术的接纳也需要经历一个过程。

上文中已经提到过，这项工作对于英美的等判例法国家具有重大的作用，在判例法国家，法官在做出裁决时首先要考虑的就是判例，通过对判例的总结与判断来做出正确的裁决。不同于判例法国家，我国对于判例并没有依赖性，对于案

件的判决完全是依照现行的法律规定，并不会依照先前存在的判例来对案件进行处理。判例在法官处理案件的过程中最多起着一个指导的作用，并不会对于案件的处理产生实质性的影响，关于判例的寻找并不是律师或者法官的工作重点，判例的作用并不是像英美法系国家那样重要。在这种情况下，TAR 技术在我国是否具有广泛适用的土壤这一点也值得探讨。

五、TAR 技术适用可能存在的法律问题

关于 TAR 技术可能涉及的法律问题在讲述 TAR 技术存在的弊端时就稍有涉及。归根到底，TAR 技术作为一种基于人工智能的文本搜索技术，目前涉及到对该技术进行规制的法律规定很少，这就使得对于该技术的应用在法律层面面临着无法可依的局面。

根据作者在本章内容中的叙述，TAR 技术在法律技术领域的运用要最终达到“彻底解放人类双手”的效果，而对于人工智能而言，其在搜索文件时并不会对于互联网上的相关文件进行预先筛选，以确定哪些文件是公开使用，哪些文件是私密的。如果不施加人为干预，可能会对网络用户的隐私权会造成很大的威胁。

其次是关于司法领域运用 TAR 技术与司法工作人员伦理的冲突问题。技术与人的理性在目标上往往具有一致性，反映到 TAR 技术上，就是尽可能全面的找到所有符合需求的文件。技术的应用在司法领域更多体现的是一种追求效益提升的效率价值，在这一层面上其与司法工作人员的伦理规范完全吻合，即勤勉工作，认真尽责。但是司法工作人员的伦理规范不单单只蕴含这一价值，除此之外，是司法工作人员在从事法律工作时所展现出来的人文关怀与情感价值。包括司法裁判在内的诸多的司法工作并不是将法律文本机械的运用于法律工作，单纯的理性思维在司法过程中并不能解决问题，其更多的是依靠司法工作人员在司法过程中对于法理与情理的考量。但是，基于 TAR 技术搜集的文本完全是按照同一标准筛选出来的，司法工作人员在对相关判例进行梳理对照时，受先前判例的影响往往会对案件形成固式化思维，从而在处理案件时忽略个案的特殊性。而在人工审查的情况下，虽然收集到的文件不会像 TAR 一样全面，但是基于人内心的情感与价值抉择，人工审查只会将与审理案件在案件事实与法理考量上最为相似的案件留下来，而不是将符合某种固式标准的案件全部留下来，从而在保证为司法人员提供判例参考的前提下避免司法人员先入为主，预先断案。同人工审查相比，TAR 技术尽管在文件的找寻上机器或许能够达到更高的精确度以及效率比值，但是民众未必能够信服其基于理性思维所做出的决定。其最终结果就是造成司法公信力的削弱。

六、关于 TAR 技术的理解与思考

首先，笔者并不赞成 TAR 技术完全实现自主化，从而进一步脱离人的干预。某种意义上这代表着文件审查这个过程中人只能被动的接受 TAR 技术的反馈结果，因此，在 TAR 技术运用的过程必须实施人为干预。其次，笔者赞成将 TAR 技术有范围的应用于司法领域，文件审查工作确实繁杂而且低效率，将 TAR 技术引入法律文件审查工作中，可以有效的提高工作效率，同时鉴于 TAR 技术推广可能导致的法律问题，必须将 TAR 技术与人工审查方式结合起来，对于基于 TAR 技术筛选出来的文件进行再审查，将与需要处理案件具有真实关联性的判例挑选出来供法律职业人员进行查阅，在为其提供参考的前提下尽可能减少对法官的影响。

上述内容就是笔者关于本章内容的一些思考与想法。总而言之，科技的进步已经对于整个社会生活产生了巨大的影响，法律领域也不可避免。除了处理某些

核心问题，较为简单的且重复性的问题交由算法处理不失为一个好办法。但是算法的应用也必须收到法律的规制，从而更好的应用于社会生活。